Pure Data for Analytics System Administration

Technical Account Management Team
For questions about this presentation contact Joe Baric at jbaricr@us.ibm.com

V 1.1

© 2013 IBM Corporation

PureData System for Analytics

System administration Agenda

- **Managing your operating system**
  - –A word on upgrades
  - –Users
  - –Variables
  - –Securely managing passwords
- **Managing the database service**
  - –Starting and stopping the system
  - –Reviewing system health
  - –Interacting with user sessions and queries
- **Interacting with system objects**
  - –Users
  - –Groups – rights and workload groups
  - –Databases
  - –Tables/views/synonyms
  - –Privileges/access rights
- **Utility scripts**
  - –Nz_toolkit scripts
  - –Other support scripts
- **Extending your environment**
  - –Installing SQL Extensions
  - –Installing IBM Netezza Analytics

© 2013 IBM Corporation

PureData System for Analytics

System Administration Agenda

- **Managing your operating system**
  - A word on upgrades
  - Users
  - Variables
  - Securely managing passwords
- **Managing the database service**
  - Starting and stopping the system
  - Reviewing system health
  - Interacting with user sessions and queries
- **Interacting with system objects**
  - Users
  - Groups – rights and workload groups
  - Databases
  - Tables/views/synonyms
  - Privileges/access rights
- **Utility scripts**
  - Nz_toolkit scripts
  - Other support scripts
- **Extending your environment**
  - Installing SQL Extensions
  - Installing IBM Netezza Analytics

© 2013 IBM Corporation

Upgrades

- IBM regularly releases software and firmware upgrades and patches. These fall into four main categories:
  - Operating system releases
  - Host management scripts (HPF)
  - System firmware (FDT)
  - Database software (NPS)

- You can subscribe for notification when these become available on the IBM website – your IBM technical team may be able to advise when patches are available as well.

- If you need advice on whether to upgrade or not reach out to your IBM account team.
  - IBM will carry out ALL upgrades – all that is needed is a PMR logged with support
  - IBM can then advise of prerequisites, timescales and effort involved
  - Currently, major upgrade activity (typically firmware) support requirese a week of notice for upgrades done between 9am – 5pm EST Mon – Fri, or two weeks outside that window, such as weekends.

‘Out of the box’ environment

- The PureData for Analytics appliance is a dedicated platform and as such is fairly simple to administer..
- The database runs on an active/passive clustered environment. There are two servers (hosts). One is the **active** host (running the database) the other is the **passive** host (waiting to take over if the active host gets in trouble)
- Each host comes with two defined users
  - root
  - nz (database power user)
- If you need more OS level users this can be done but you will need to actively manage these users yourself and be aware of implications of the active/passive cluster environment
- The filesystems are standard, pre-allocated and cannot be changed. The /nz and /export/home filesystems are part of the cluster and in case of failover from the active to passive host will migrate.
- Therefore if you have processes that rely on scripts or data being available it’s recommended these are placed in /export/home/ (possibly in the nz user homes directory, /export/home/nz)

Environment Variables

Variable	Notes
NZ_HOST	Set to the hostname or IP address of the NPS System
NZ_DATABASE	Set to the specific database to connect to
NZ_USER	The database user you wish to connect as
NZ_PASSWORD	Authentication password for the database user

- There are several environment variables required to support the NPS database administration and system administration tools:
  - NZ_HOST, NZ_DATABASE, NZ_USER are the most commonly used environment variables and by default will be set in the user's .bashrc file
  - If these are not set on the active host you will be prompted for the details every time you call a system command.
- To avoid displaying the password in the environment variable use the nzpassword command to create a locally stored encrypted password.
- These variables apply for any client tools on other boxes (such as ETL platforms).

PureData System for Analytics

System Administration Agenda

- **Managing your operating system**
  - A word on upgrades
  - Users
  - Variables
  - Securely managing passwords
- **Managing the database service**
  - Starting and stopping the system
  - Reviewing system health
  - Interacting with user sessions and queries
- **Interacting with system objects**
  - Users
  - Groups – rights and workload groups
  - Databases
  - Tables/views/synonyms
  - Privileges/access rights
- **Utility scripts**
  - Nz_toolkit scripts
  - Other support scripts
- **Extending your environment**
  - Installing SQL Extensions
  - Installing IBM Netezza Analytics

© 2013 IBM Corporation

Starting and stopping the system

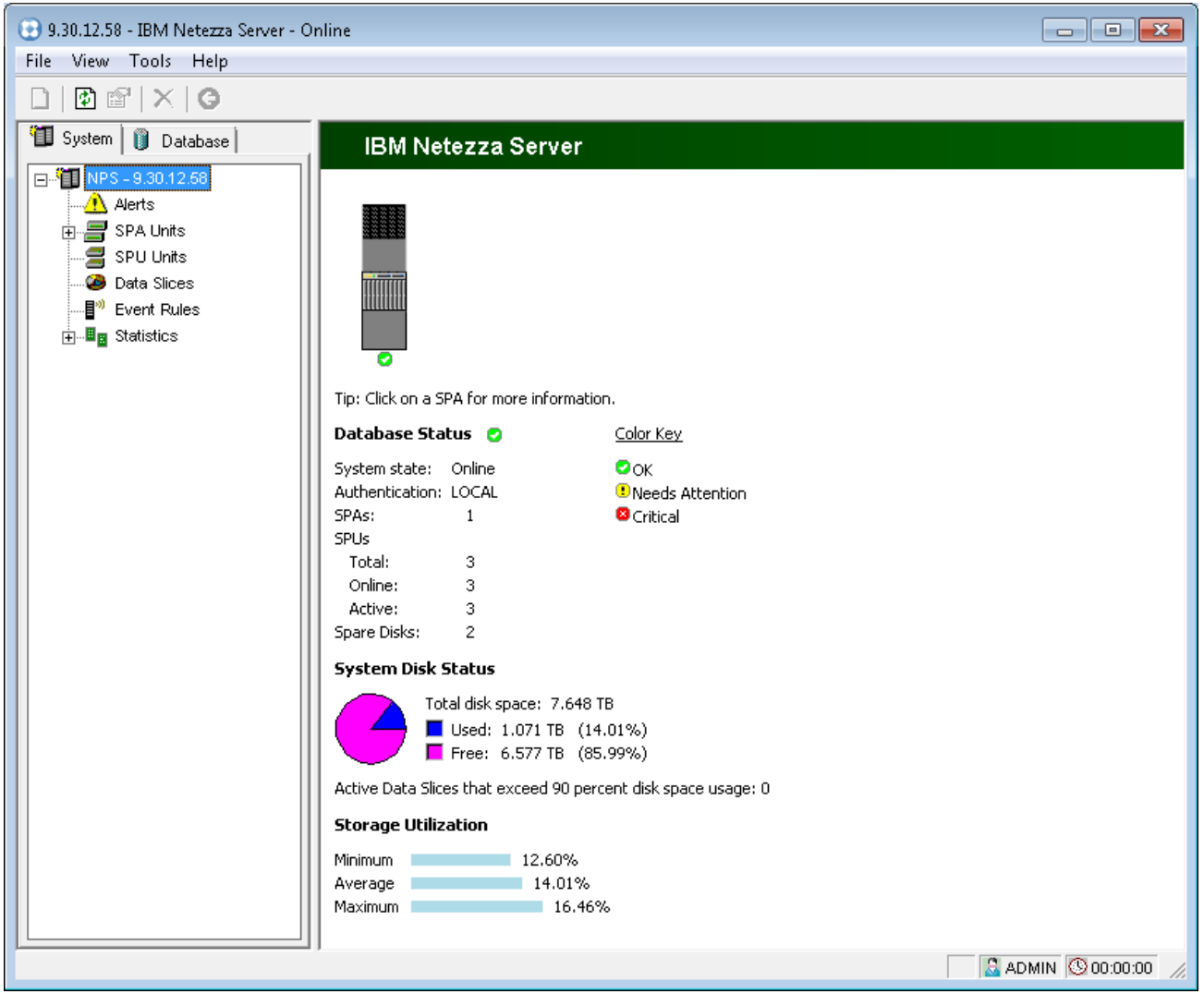
- `nzstart <options>`
  - Command used to start the NPS System.
  - Same as ‘nzsystem start’
- `nzstop <options>`
  - Command used to stop the NPS System.
  - Same as ‘nzsystem stop’

- **The `nzstart` and `nzstop` commands must be run from the NPS host as the user nz.**
- `nzsystem restart`
  - Command used to stop and start the database immediately

System state changes

- When the PureData appliance starts, or when nzstart is run, the system goes through a number of state changes:
  - `Stopped` → `Initializing` → `Pre-Online` → `Resuming` → `Online`
    - `Stopped` — The system is not running
    - `Initializing` — The system is transitioning to initialized
    - `Pre-Online` — The first phase of `Resuming`
    - `Resuming` — The system is transitioning to `Online`
    - `Online` — System is running normally
- `nzstate show`
  - Command to display the current state of the NPS System
- Don't worry about intermediate states. The main thing to focus on is if the system is online or not.

System state changes



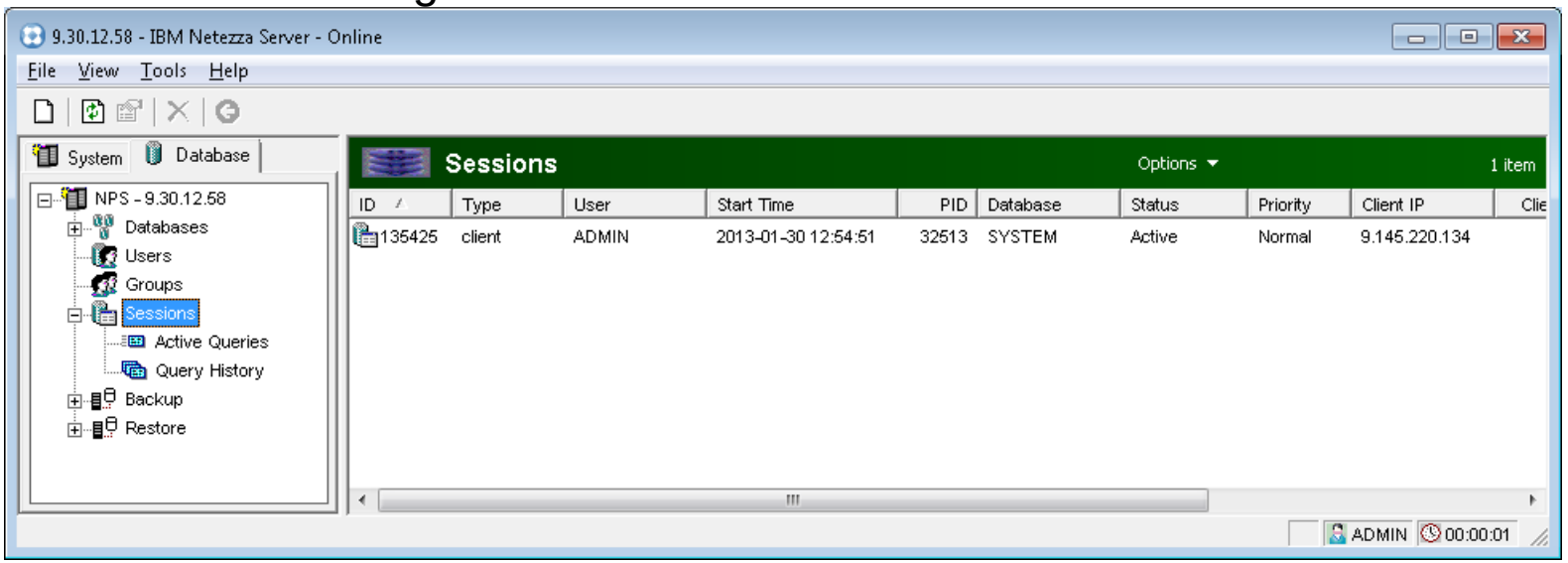
Viewing connected sessions

- Command line – nzsession

```
- [nz@nz80409-h1 ~]$ nzsession

- ID    Type User  Start Time              PID   Database State  Priority Name Client IP      Client PID Command
- ----- ---- ----- ----------------------- ----- -------- ------ ---------- -------------- ---------- ------------------------
- 74911 sql  ADMIN 21-May-12, 11:09:52 EDT 13492 SYSTEM   idle   normal        172.16.251.190       4744 select ds.ds_id, prm.sdm
- 74934 sql  ADMIN 21-May-12, 11:12:17 EDT 14202 SYSTEM   active normal             127.0.0.1      14201 SELECT session_id, clien
```

- GUI
  - Needs refreshing to see the current state



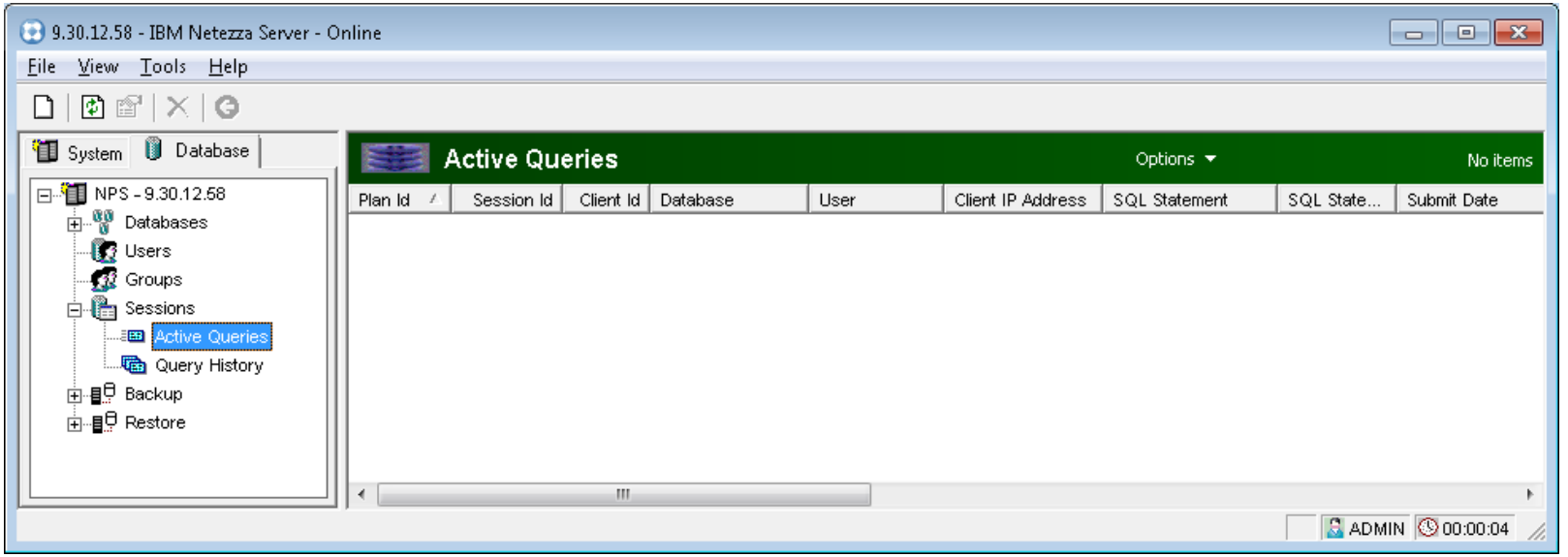
Viewing active queries

- Command line – nzsession -activetxn

```
[nz@nz80409-h1 ~]$ nzsession -activetxn

ID    Type User  Start Time              PID   Database State  Priority Name Client IP      Client PID Command
----- ---- ----- ----------------------- ----- -------- ------ ------------ -------------- ---------- -----------------------
74934 sql  ADMIN 21-May-12, 11:12:17 EDT 14202 SYSTEM   active normal             127.0.0.1      14201 SELECT session_id, clien
```

- GUI
  - Needs refreshing to see the current state



Aborting sessions

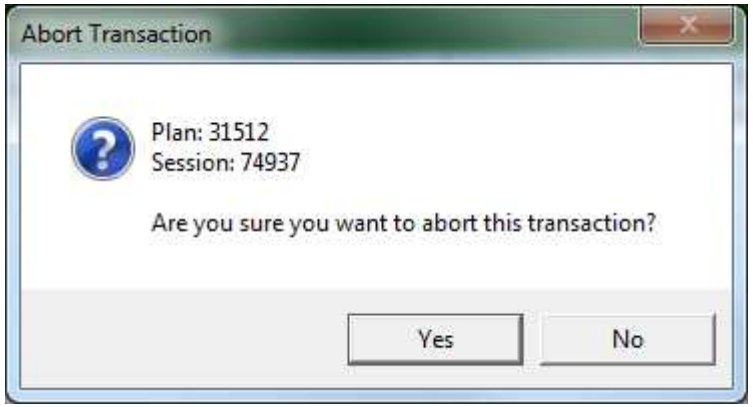
- Command line – nzsession

```
- [nz@nz80409-h1 ~]$ nzsession

- ID    Type User  Start Time              PID   Database State  Priority Name Client IP      Client PID Command
- ----- ---- ----- ----------------------- ----- -------- ------ ------------ -------------- ---------- ------------------------
- 74911 sql  ADMIN 21-May-12, 11:09:52 EDT 13492 SYSTEM   idle   normal        172.16.251.190       4744 select q.qs_planid, q.qs
- 74956 sql  ADMIN 21-May-12, 11:15:05 EDT 15112 SYSTEM   active normal             127.0.0.1      15111 SELECT session_id, clien

- [nz@nz80409-h1 ~]$ nzsession abort -id 74956
- Are you sure you want to abort the session (y|n)? [n] y
```

- GUI
- Right click > Abort



PureData System for Analytics

System Administration Agenda

- **Managing your operating system**
  - A word on upgrades
  - Users
  - Variables
  - Securely managing passwords
- **Managing the database service**
  - Starting and stopping the system
  - Reviewing system health
  - Interacting with user sessions and queries
- **Interacting with system objects**
  - Users
  - Groups – rights and workload groups
  - Databases
  - Tables/views/synonyms
  - Privileges/access rights
- **Utility scripts**
  - Nz_toolkit scripts
  - Other support scripts
- **Extending your environment**
  - Installing SQL Extensions
  - Installing IBM Netezza Analytics

© 2013 IBM Corporation

Users

- Standard SQL commands are all documented in the Database User guide:
- Create User:
  - `CREATE USER test_user W?IH PASSWORD ‘test_user’`
- Change user password:
  - `ALTER USER test_user WITH PASSWORD ‘new password’`
- Change user password and pre-expire:
  - `ALTER USRE test_user WITH PASSWORD ‘new password’ EXPIRE PASSWORD:`
- Add user to a resource group (for workload management):
  - `ALTER USER test_user WITH RESOURCEGROUP resource_group1?`
- Add to a rights group (for privilege management)
  - `ALTER USER test_user WITH IN GROUP rights_group1`
- Drop a user (who must not own any objects – we can’t cascade drop)
  - `DROP USER test_user;`
  - You can do all this via the admin GUI. **However** SQL is easier to record, play back and repeat as needed.

Users - CLI

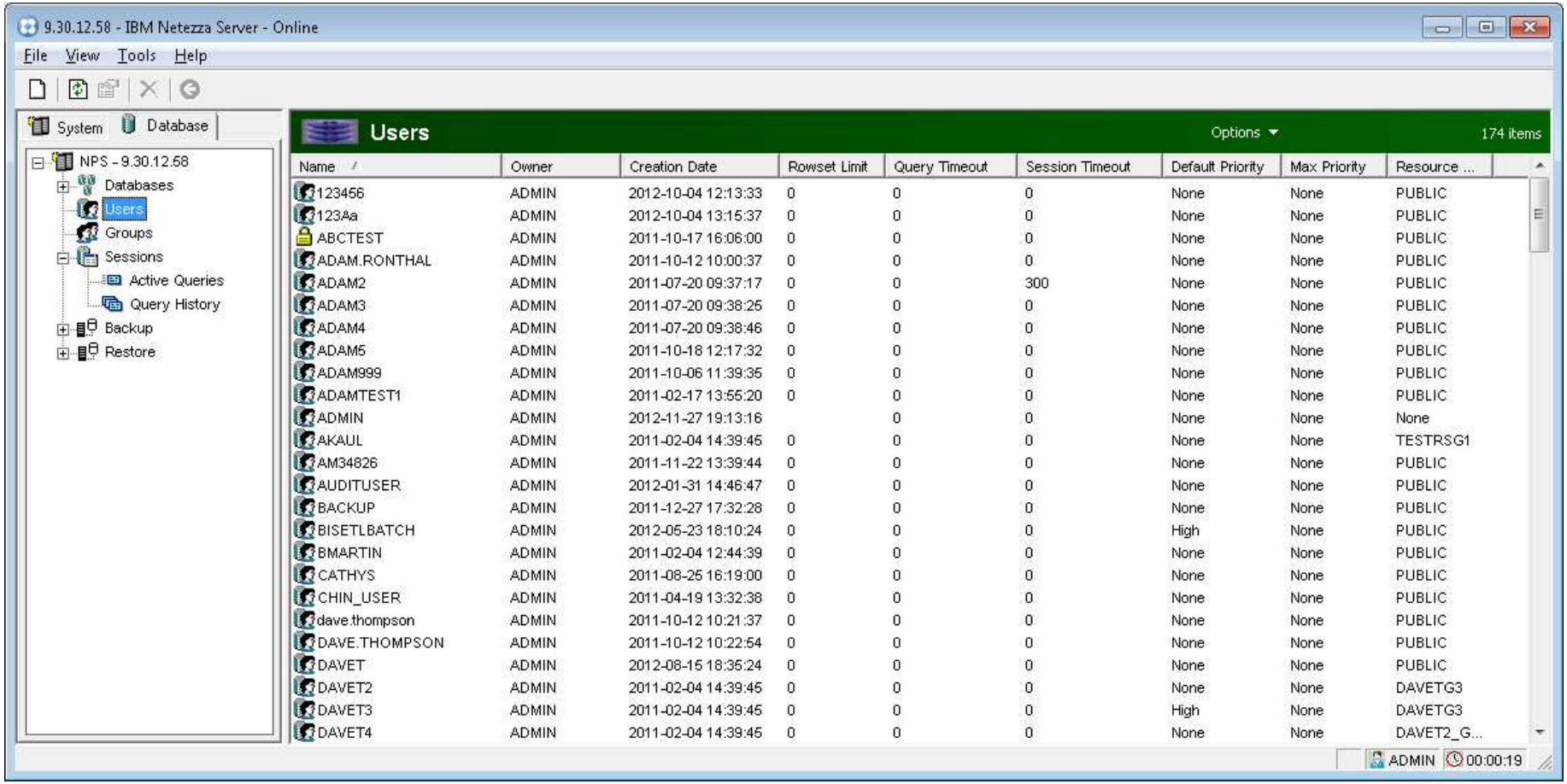
- List users:

```
[nz@nz80409-h1 ~]$ nzsql
Welcome to nzsql, the IBM Netezza SQL interactive terminal.

Type:  \h for help with SQL commands
       \? for help on internal slash commands
       \g or terminate with semicolon to execute query
       \q to quit

SYSTEM(ADMIN)=> \du
                                                                List of Users
     USERNAME      |     VALIDUNTIL      | ROWLIMIT | SESSIONTIMEOUT | QUERYTIMEOUT | DEF_PRIORITY | MAX_PRIORITY | USERESOURCEGRPID | USERESOURCEGRPNAME |
   CROSS_JOINS_ALLOWED
-------------------+---------------------+----------+----------------+--------------+--------------+--------------+------------------+--------------------+------------
   ---------
 ABCTEST           |                     |        0 |              0 |            0 | NONE         | NONE         |                  | PUBLIC             | NULL
 ADAM.RONTHAL      |                     |        0 |              0 |            0 | NONE         | NONE         |                  | PUBLIC             | NULL
 ADAM2             |                     |        0 |            300 |            0 | NONE         | NONE         |                  | PUBLIC             | NULL
 ADAM3             |                     |        0 |              0 |            0 | NONE         | NONE         |                  | PUBLIC             | NULL
 ADAM4             |                     |        0 |              0 |            0 | NONE         | NONE         |                  | PUBLIC             | NULL
 ADAM5             |                     |        0 |              0 |            0 | NONE         | NONE         |                  | PUBLIC             | NULL
 ADAM999           |                     |        0 |              0 |            0 | NONE         | NONE         |                  | PUBLIC             | NULL
 ADAMTEST1         |                     |        0 |              0 |            0 | NONE         | NONE         |                  | PUBLIC             | NULL
```

Users - GUI


9.30.12.58 - IBM Netezza Server - Online
File View Tools Help
System Database
NPS - 9.30.12.58
Databases
Users
Groups
Sessions
Active Queries
Query History
Backup
Restore
Users
Options
174 items
Name | Owner | Creation Date | Rowset Limit | Query Timeout | Session Timeout | Default Priority | Max Priority | Resource ...
123456 | ADMIN | 2012-10-04 12:13:33 | 0 | 0 | 0 | None | None | PUBLIC
123Aa | ADMIN | 2012-10-04 13:15:37 | 0 | 0 | 0 | None | None | PUBLIC
ABCTEST | ADMIN | 2011-10-17 16:06:00 | 0 | 0 | 0 | None | None | PUBLIC
ADAM.RONTHAL | ADMIN | 2011-10-12 10:00:37 | 0 | 0 | 0 | None | None | PUBLIC
ADAM2 | ADMIN | 2011-07-20 09:37:17 | 0 | 0 | 300 | None | None | PUBLIC
ADAM3 | ADMIN | 2011-07-20 09:38:25 | 0 | 0 | 0 | None | None | PUBLIC
ADAM4 | ADMIN | 2011-07-20 09:38:46 | 0 | 0 | 0 | None | None | PUBLIC
ADAM5 | ADMIN | 2011-10-18 12:17:32 | 0 | 0 | 0 | None | None | PUBLIC
ADAM999 | ADMIN | 2011-10-06 11:39:35 | 0 | 0 | 0 | None | None | PUBLIC
ADAMTEST1 | ADMIN | 2011-02-17 13:55:20 | 0 | 0 | 0 | None | None | PUBLIC
ADMIN | ADMIN | 2012-11-27 19:13:16 | | 0 | 0 | None | None | None
AKAUL | ADMIN | 2011-02-04 14:39:45 | 0 | 0 | 0 | None | None | TESTRSG1
AM34826 | ADMIN | 2011-11-22 13:39:44 | 0 | 0 | 0 | None | None | PUBLIC
AUDITUSER | ADMIN | 2012-01-31 14:46:47 | 0 | 0 | 0 | None | None | PUBLIC
BACKUP | ADMIN | 2011-12-27 17:32:28 | 0 | 0 | 0 | None | None | PUBLIC
BISETLBATCH | ADMIN | 2012-05-23 18:10:24 | 0 | 0 | 0 | High | None | PUBLIC
BMARTIN | ADMIN | 2011-02-04 12:44:39 | 0 | 0 | 0 | None | None | PUBLIC
CATHYS | ADMIN | 2011-08-25 16:19:00 | 0 | 0 | 0 | None | None | PUBLIC
CHIN_USER | ADMIN | 2011-04-19 13:32:38 | 0 | 0 | 0 | None | None | PUBLIC
dave.thompson | ADMIN | 2011-10-12 10:21:37 | 0 | 0 | 0 | None | None | PUBLIC
DAVE.THOMPSON | ADMIN | 2011-10-12 10:22:54 | 0 | 0 | 0 | None | None | PUBLIC
DAVET | ADMIN | 2012-08-15 18:35:24 | 0 | 0 | 0 | None | None | PUBLIC
DAVET2 | ADMIN | 2011-02-04 14:39:45 | 0 | 0 | 0 | None | None | DAVETG3
DAVET3 | ADMIN | 2011-02-04 14:39:45 | 0 | 0 | 0 | High | None | DAVETG3
DAVET4 | ADMIN | 2011-02-04 14:39:45 | 0 | 0 | 0 | None | None | DAVET2_G...
ADMIN
00:00:19

Databases

- A PureData for Analytics database is similar to (but not the same as!) a schema in other relational database products – it's a logical construct.
- This means it is quick to create/change and there are few options.
- Create
  - `CREATE DATABASE new_database`
- Rename
  - `ALTER DATABASE new_database RENAME old_database`
- When creating databases consider:
  - Size of the database (backups are at the database level)
  - Permissions (it's easier to manage permissions at the database level so consider grouping data that is going to be accessed by common groups of users and applications)

Databases (pre 7.0.3)

- You can select from tables in databases you're not connected to.
- There is a 'placeholder' for schema in the naming convention. You can fill this with anything you want.. Or nothing. The following are functionally the same when referring to a table in a database:
  - `database.schema.table`
  - `database.admin.table`
  - `database.cheese.table`
  - `Database..table`

```
SELECT count(*)
FROM database1..table1 t1
INNER JOIN
     database2..table2 t2
ON t1.col1 = t2.col2;
```

- You can only make changes to objects and data in the database you're connected to

```
DB1(ADMIN)=> insert into DB2..test_table values(1);
ERROR:  Cross Database Access not supported for this type of command
```

Tables

- Netezza DDL is simpler than many other systems. Minimum you should specify:
  - Table name
  - Column names and datatypes
  - Distribution key

```
CREATE TABLE stuart_table (col1 INTEGER
                          ,col2 VARCHAR(100)
                          ,col3 DATE)
DISTRIBUTE ON (col1);
```

- Always specify distribution key
- Optionally you can specify:
  - PK/FK relationships (although these are not enforced!)
  - NOT NULL
  - Default values
  - ORGANIZE ON columns for clustered base tables

Tables

- Tables are owned by the user that created them but ownership can be passed on.
  - ALTER TABLE stuart_table OWNER TO new_user
- You can rename columns, add/drop columns
  - But consider creating a new copy of the table instead.
  - If you add or drop a column the system creates a NEW version of the table. As data is written or updated it is put in this new table copy.
  - Over time, performance will degrade as the system effectively UNIONs these two versions of the table.
  - Solution is to run GROOM TABLE tablename VERSIONS as soon as possible after the alter table.
- See the user guide for additional commands.

Views

- Create syntax is as you would expect:

  CREATE VIEW stuart_view AS

  SELECT * FROM stuart_table

- You can select from tables in other databases

  CREATE VIEW stuart_view AS

  SELECT * FROM database2..stuart_table

- If an object a view uses changes, the view will need to be recreated.

Synonyms

- Use synonyms where:
  - You want to reference an object by a different name
  - You want to reference an object in another database without using the database qualifier
- You can use a synonym on a table, view, aggregate or function (very useful for user defined functions and aggregates– this will be revisited later).

Other object types

- PureData for Analytics systems support a number of object types. We have not discussed:
  - Stored procedures
  - Materialized views
  - Sequences
  - External tables
  - Temporary tables
- These are all available and discussed in depth in the user manuals and training courses available.

PureData System for Analytics

System Administration Agenda

- **Managing your operating system**
  - A word on upgrades
  - Users
  - Variables
  - Securely managing passwords
- **Managing the database service**
  - Starting and stopping the system
  - Reviewing system health
  - Interacting with user sessions and queries
- **Interacting with system objects**
  - Users
  - Groups – rights and workload groups
  - Databases
  - Tables/views/synonyms
  - Privileges/access rights
- **Utility scripts**
  - Nz_toolkit scripts
  - Other support scripts
- **Extending your environment**
  - Installing SQL Extensions
  - Installing IBM Netezza Analytics

© 2013 IBM Corporation

Support toolkit scripts

- In /nz/support/bin you will find a large library of very useful scripts.
- You will also find a readme.html file. Download to your PC! This file contains the help text for over 250 utility scripts designed to make your life easier.
- If you want a one-liner on each script's purpose.. there is a script for that!

  nz_help –h

  Usage: nz_help [search_string]

  Purpose: Provide help (a listing) of the scripts that are included with this toolkit.

- Each script has help built into it.

  nz_scriptname -?

Other supporting scripts

- /nz/support/bin/adm/disk_monitor
  - Logs information on disk health and recommends proactive action on disks that may be failing
- /nz/kit/bin/adm/nzhealthcheck
  - Will review system health and can report on a variety of issues.

PureData System for Analytics

System administration Agenda

- **Managing your operating system**
  - A word on upgrades
  - Users
  - Variables
  - Securely managing passwords
- **Managing the database service**
  - Starting and stopping the system
  - Reviewing system health
  - Interacting with user sessions and queries
- **Interacting with system objects**
  - Users
  - Groups – rights and workload groups
  - Databases
  - Tables/views/synonyms
  - Privileges/accessrights
- **Utility scripts**
  - Nz_toolkit scripts
  - Other support scripts
- **Extending your environment**
  - Installing SQL Extensions
  - Installing IBM Netezza Analytics

© 2013 IBM Corporation

SQL Extensions

- PureData for Analytics appliances have a large number of built-in functions that are typically ANSI standard functions. In addition we provide the functionality for customers to create their own functions and aggregates.
- This has been used by IBM to provide a wide array of fully supported functions that expand the flexibility of the PureData platform. These include:
  - Array processing
  - Regular expressions
  - XML manipulation
  - Text encoding, compression and manipulation
  - Date manipulation

SQL Extensions

- Can be downloaded from FixCentral
- The install package includes a user manual and installation guide
- When installing, it is advisable to install the functions in a single central database.
- From here you can create synonyms in all databases that require this functionality
  - This means that upgrading SQL Extensions only has to take place in one database.
  - Access can be easily controlled.

- **Installing or uninstalling SQL Extensions does not require system downtime.**

- Example: Assume the SQL Extensions have been unpacked to /export/home/nz/sql_toolkit/
- Create a database to hold the functions:

```
create database sql_ext;
```

- Register the functions:

```
/export/home/nz/sql_toolkit/libnetcrypto/1.5/install-d sql_ext

/export/home/nz/sql_toolkit/libnetxml/1.5/install-d sql_ext
```

- Generate CREATE SYNONYM commands to run in databases you want to use the SQL Extensions in:

```
select distinct 'create synonym '||function||' for sql_ext..'||function||';'
  from _v_function where builtin = ‘f';

select distinct 'create synonym '||aggregate||' for sql_ext..'||aggregate||';' from
  _v_aggregate where builtin = ‘f';
```

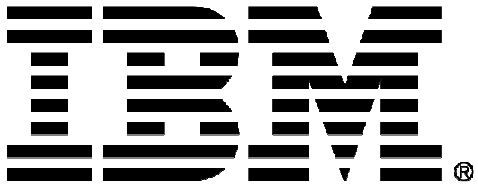
- Run these generate statements when a database needs these functions available.
- Ensure users have the correct rights on these functions and aggregates:

```
Grant execute on function,aggregate to new_user;
```

IBM Netezza Analytics

- The PureData for Analytics was built with advanced analytics and data processing at its core
- In order to facilitate this, IBM provides a comprehensive collection of analytic functionality designed from the ground up to work on an AMPP platform
- Details on the functionality provided is available on request!

- In order to install this functionality, download the software package from FixCentral.
- An installation guide is included and should be followed. The installation does not require downtime.
- The INZA functions will by default be installed in a central location and should be referenced as required.



© International Business Machines Corporation 2012
International Business Machines Corporation New Orchard Road Armonk, NY 10504
IBM, the IBM logo, PureSystems, PureFlex, PureApplication, PureData and ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide.
A current list of IBM trademarks is available on the Web at www.ibm.com/legal/copytrade.shtml
All rights reserved.

WAP12402-USEN-01